

RESEARCH

Open Access



Construction and validation of a prognostic nomogram for predicting the survival of HIV/AIDS adults who received antiretroviral therapy: a cohort between 2003 and 2019 in Nanjing

Fangfang Jiang^{1†}, Yuanyuan Xu^{2†}, Li Liu², Kai Wang³, Lu Wang⁴, Gengfeng Fu⁵, Liping Wang⁶, Zhongjie Li⁶, Junjie Xu⁷, Hui Xing⁴, Ning Wang⁴, Zhengping Zhu^{2*} and Zhihang Peng^{1*}

Abstract

Background: Great achievements have been achieved by free antiretroviral therapy (ART). A rapid and accurate prediction of survival in people living with HIV/AIDS (PLHIV) is needed for effective management. We aimed to establish an effective prognostic model to forecast the survival of PLHIV after ART.

Methods: The participants were enrolled from a follow-up cohort over 2003–2019 in Nanjing AIDS Prevention and Control Information System. A nested case-control study was employed with HIV-related death, and a propensity-score matching (PSM) approach was applied in a ratio of 1:4 to allocate the patients. Univariable and multivariable Cox proportional hazards analyses were performed based on the training set to determine the risk factors. The discrimination was qualified using the area under the curve (AUC) and concordance index (C-Index). The nomogram was calibrated using the calibration curve. The clinical benefit of prognostic nomogram was assessed by decision curve analysis (DCA).

Results: Predictive factors including CD4 cell count (CD4), body mass index (BMI) and hemoglobin (HB) were determined and incorporated into the nomogram. In the training set, AUC and C-index (95% CI) were 0.831 and 0.798 (0.758, 0.839), respectively. The validation set revealed a good discrimination with an AUC of 0.802 and a C-index (95% CI) of 0.786 (0.681, 0.892). The calibration curve also exhibited a high consistency in the predictive power (especially in the first 3 years after ART initiation) of the nomogram. Moreover, DCA demonstrated that the nomogram was clinically beneficial.

Conclusion: The nomogram is effective and accurate in forecasting the survival of PLHIV, and beneficial for medical workers in health administration.

Keywords: HIV/AIDS, Antiretroviral therapy, Prognostic model, Nomogram

Introduction

Over the past 30 years, HIV has become a major global public health challenge [1]. In China, free antiretroviral therapy (ART), launched in 2003, has proven to efficiently recover CD4 cell count (CD4), lower viral load (VL) and curb HIV transmission [2]. Nevertheless, the poor prognosis of people living with HIV/AIDS (PLHIV) after ART

*Correspondence: zzp@njcdc.cn; zhihangpeng@njmu.edu.cn

†Fangfang Jiang and Yuanyuan Xu contributed equally to this work.

¹ Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 211166, Jiangsu, China

² Department of AIDS and STDs control and prevention, Nanjing Municipal Center for Disease Control and Prevention, Nanjing 210003, Jiangsu, China

Full list of author information is available at the end of the article



remains a concern [3, 4]. It is essential to create a tool to rapidly and accurately predict death risk among PLHIV.

Studies have shown that CD4, CD8 cell count (CD8), and VL before treatment are closely associated with the mortality of PLHIV [5–13]. Clinical indicators are reported to have a close association with death risk of PLHIV [5, 7–16]. Some laboratory indicators, such as hemoglobin (HB), platelet-related indexes, are also related to the progression and mortality of HIV-related diseases after ART [3, 17–23].

Since the combination of several independent indicators, rather than a single predictive factor, has a stronger predictive power, several scoring systems based on the multiple risk factors have been proposed to forecast the mortality of PLHIV. However, there still lacks a widely-held effective scoring system to predict the survival of ART-treated PLHIV.

In recent years, various multi-factor models have been designed to estimate disease outcomes. A risk-scoring system can be established according to the recommendation of Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) [24]. Nomogram is convenient to predict the prognosis of patients [25]. Previous nomograms have failed to assess the outcomes of ART. For example, in the model by Margaret et al. [26], a concordance index (C-Index) of 0.75 (95% CI: 0.74–0.81) in the training set and a C-Index of 0.69 (95% CI: 0.59–0.77) in the validation set were presented. This model achieves a satisfactory performance, but far from excellent. Few prognostic models based on PLHIV after receiving ART have presented good discrimination and calibration. In the model established by Hou et al. [27], the C-Indexes are 0.91 (95% CI: 0.86–0.97) in the training set and 0.92 (95% CI: 0.82–1.00) in the validation set.

In the present study, to build a simple and effective prognostic model to forecast the survival of PLHIV after ART, a nested case-control study was employed with HIV-related mortality events, and a propensity-score matching (PSM) approach was applied to allocate the patients in a ratio of 1:4. To make the model more reliable and robust, bootstrap was used for internal validation. The discrimination and calibration of the model were evaluated based on the training set and validation set. Decision curve analysis (DCA) was also used to evaluate the performance of the nomogram.

Materials and methods

Study design

The data used in this study were extracted from patients who received ART between 2003 and 2019 from Nanjing AIDS Prevention and Control Information System (AIDS-PCIS). All patients received a free ART containing

at least three antiviral medicines. The follow-up started after ART initiation and the participants were visited every 3 months. The observation end point was December 31, 2019, and the outcome was death. The survival time was defined as the duration from ART initiation to death or December 31, 2019. The inclusion criteria included: (1) living in Nanjing; (2) being visited at least once; (3) being over 18 years old when ART started; and (4) having complete laboratory test data before starting ART. At the end of the follow-up, a total of 4573 patients met the inclusion criteria. Among them, 120 patients died of HIV/AIDS-related diseases and were determined as the cases in the nested case-control study. The flow-chart of recruitment participants is shown in the Fig. 1.

Data collection

Demographic data and clinical information were retrieved from face-to-face surveys at the patients' enrollment or extracted from their medical records using a structured questionnaire designed specifically for AIDS-PCIS. The information included the date of birth, gender, height, weight, marital status, infection route and WHO clinical stage. The age of the patient was calculated from the date of birth to the date of starting ART. Body mass index (BMI) was calculated using the following formula: $BMI = \text{weight (kg)} / (\text{height (m)} \times \text{height (m)})$.

The laboratory testing data were obtained from the Nanjing Center for Disease Control and Prevention (CDC) or local hospitals. The laboratory testing indicators included CD4, white blood cell (WBC), blood platelets (PLT), HB, serum creatinine (CR), triglycerides (TG), total cholesterol (TC), fasting blood glucose (FBG), aspartate aminotransferase (AST), alanine aminotransferase (ALT) and total bilirubin (TBIL). All these laboratory tests were carried out by the trained technical personnel strictly following clinical guidelines at each visit in the central laboratory of local hospitals or Nanjing CDC.

Routine blood biochemical indexes, such as TG, TC, FBG, CR, AST, ALT, and TBIL, were measured using a Beckman AU5800 automatic biochemical analyzer (Beckman COULTER K., Japan). Other indexes including WBC, HB and PLT were evaluated by Sysmex Xe-2100 automatic blood cell analyzer (Sysmex Corporation, Japan). CD4 was determined by the BD FACSCalibur flow cytometer (Becton Dickinson Corporation, USA).

Statistical analysis

Data processing

For a multi-factor regression model, there is no simple method to estimate its proper sample size. When the number of predictors is much larger than that of outcomes, overfitting may occur. Previous literature showed

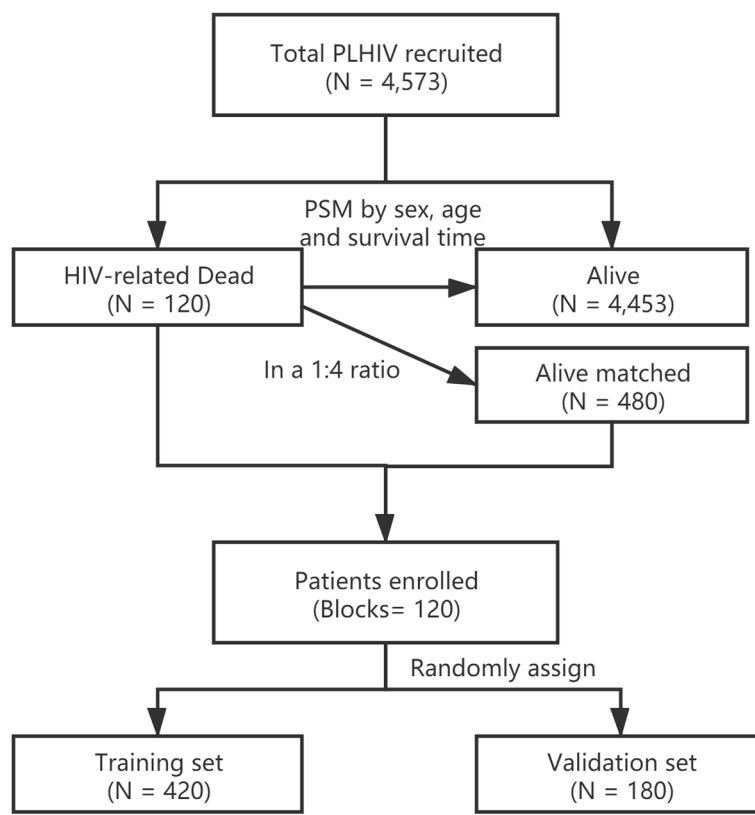


Fig. 1 A flowchart of predicted HIV-related survival of people living with HIV/AIDS (PLHIV) using nomogram model

that in the conservative estimation, one prediction factor requires at least 10 effective outcomes. In this study, there were 120 cases with effective outcomes, so the number of predictors should be less than 12.

Since directly dropping the data with missing values might lead to selection bias, or decrease the power of a test, missing value imputation was applied to obtain suitable values by employing the values of other variables before data analysis. The results were listed in Fig. 2. A sensitivity analysis was carried out to evaluate the filling effect of the missing values (Table 1).

A total of 120 deaths caused by HIV/AIDS-related diseases were determined as the cases in the nested case-control study. S(60) was set as the index date (month). To ensure that all the subjects in the case group could have a matching control, PSM was applied in a ratio of 1:4 to determine the participants (a case was well matched with 4 controls in age, gender and index date) [28]. Finally, 600 subjects were included in this study with 120 dead and 480 alive PLHIV who were separated into 120 blocks.

Establishment and validation of prediction model

The patients were randomly split into a training set and a validation set in a ratio of 7:3. The comparability of the

training set and validation set was then evaluated. Continuous variables with normal distribution were presented as mean ± standard deviation, and t-tests were used to infer the differences between the training and validation sets. The continuous variables with skewed distribution were described using median (first quartile, second quartile). The Wilcoxon rank-sum tests were employed for comparisons. Frequency (ratio) was utilized to describe the characteristics of categorical variables, and comparisons between the two sets were performed using chi-square tests or Fisher’s exact tests.

Then the data in the training set were used to fit a model and the data in validation set were applied to evaluate the efficacy of the model. Based on the data in the training set, univariable Cox proportional hazards analysis was performed for each variable. P-values of the variables were calculated based on the univariable Cox proportional hazards regression model. The variables with p-values less than or equal to 0.2 were included in a multivariable Cox proportional hazards regression model. After the multivariable analysis, the factors with p-value less than or equal to 0.05 were included in the prediction model. According to Occam’s Razor, the model with the fewest variables is the best [29]. Finally,

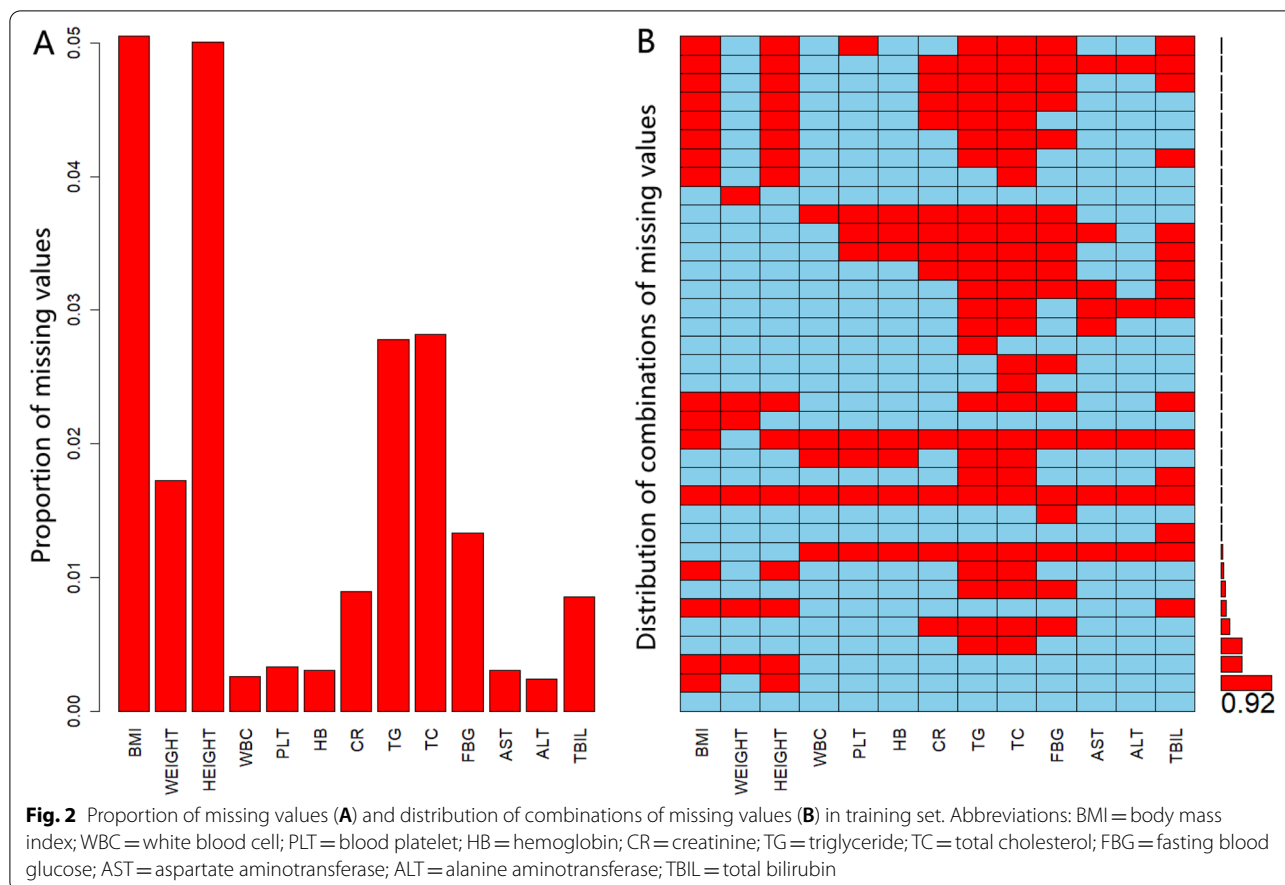


Table 1 Sensitivity analysis in imputation for missing data

Variables	Before imputation(mean ± standard deviation)	After imputation(mean ± standard deviation)	P value
Weight, kg	64.8 ± 10.2	64.8 ± 10.1	0.953
Height, cm	172.0 ± 6.1	172.0 ± 6.1	0.828
BMI, kg/m ²	21.9 ± 3.0	21.8 ± 3.0	0.564
WBC, 10 ⁹ /L	5.7 ± 2.0	5.7 ± 2.0	0.991
PLT, 10 ⁹ /L	188.7 ± 62.5	188.6 ± 62.5	0.961
HB, g/L	142.7 ± 30.9	142.7 ± 30.9	0.995
CR, mmol/L	71.3 ± 19.4	71.3 ± 19.4	0.932
TG, mmol/L	1.7 ± 2.3	1.7 ± 2.3	0.928
TC, mmol/L	4.3 ± 1.3	4.3 ± 1.3	0.976
FBG, mmol/L	5.6 ± 1.3	5.6 ± 1.3	0.922
AST, U/L	27.0 ± 27.2	27.1 ± 27.2	0.955
ALT, U/L	30.2 ± 29.2	30.2 ± 29.2	0.986
TBIL, mmol/L	12.0 ± 6.2	12.0 ± 6.2	0.964

Abbreviations: BMI body mass index, WBC white blood cell, PLT blood platelet, HB hemoglobin, CR creatinine, TG triglyceride, TC total cholesterol, FBG fasting blood glucose, AST aspartate aminotransferase, ALT alanine aminotransferase, TBIL total bilirubin

we considered both the statistically significant risk factors and professionally significant factors, such as the difficulty of index measurement, the cost of measurement

and the difficulty of application, and then determined the predictive factors and select a prediction model with the best predictive performance.

The repeatability and extrapolation of the prediction model should be evaluated. A strict evaluation of the prediction model should include internal validation and external validation. The internal validation is performed using the same dataset as the training set. This study employed the bootstrap resampling [30] for internal validation because of the lack of additional data to verify the model. The 1000 resampling performances of the model were averaged as the internal validation performance.

Discrimination and calibration are the two most common evaluation indicators. The discrimination of the prediction model is quantified using the area under the curve (AUC) and C-Index. The C-Index value ranges from 0 to 1. The closer C-Index is to 1, the better the discrimination of the model is. A C-Index of 0.5 indicates that the model has no predictive ability. When C-Index is less than 0.5, the model prediction is contrary to the actual results. In general, a C-Index of 0.7 indicates a good prediction performance of the model. However, discrimination cannot reflect whether the estimate of absolute risk of prediction model is accurate or not because it is only based on risk scores or the ranking of prediction probabilities. Calibration is a more accurate indicator to qualify the prediction model. In this study, the calibration of the model was evaluated using the calibration curve. We sorted the predicted probabilities of all participants from the smallest to the largest, and divided the patients into ten equal parts. The average predicted probability of patients in each divided part was used as x-axis and the proportion of actual events as y-axis. Ideally, the calibration graph was a straight line with an intercept of 0 and a slope of 1. The predictive ability of the model was also evaluated using decision curve analysis (DCA).

Integrated discrimination improvement (IDI), net reclassification index or improvement (NRI) and other indicators that are used to compare models or evaluate the increase in predictive performance of individual predictors were not discussed in the present study.

Presentation of nomogram

The prediction model was visualized and presented by a nomogram. To calculate the score of each variable at each level, a scoring standard was developed based on the standard regression coefficients of all variables. Then using the scores of these factors, we calculated a total score to indicate the survival probability of each patient.

All data analyses and figures were made using R software version 4.1.0. All hypothesis tests were two-sided, with an α level of 0.05.

Results

Establishment of prediction model

In this PSM-based nested case-control study, the characteristics of the 600 PLHIV (420 from the training set and 180 from the validation set) revealed that both sets were similar in all variables (Table 2).

In the univariable Cox proportional hazards regression analysis of the training set, infection route, baseline Tuberculosis (TB), continuous diarrhea, continuous or intermittent fever, shingles, WHO clinical stage, CD4, BMI, HB, CR, TC, FBG, AST and ALT were detected to be statistically related to the mortality of PLHIV (Table 3). Variables with p -value less than or equal to 0.2 in the univariable analysis were included in the multivariable Cox proportional hazards regression model. To avoid multicollinearity caused by the strong relationship between WHO clinical stage and CD4, WHO clinical stage was not included in the multivariable Cox proportional hazards regression model. Shingles, CD4, BMI, HB and TC were found linked to HIV/AIDS-related death. In order to establish an optimal prediction model, the individual and combined performance of these factors were then evaluated using ROC analysis and C-Index. As shown in Fig. 3A, the AUCs of Shingles, CD4, BMI, HB and TC in the training set were 0.549, 0.755, 0.729, 0.669 and 0.596, respectively. The AUC of combine 1 (Shingles + CD4 + BMI + HB + TC) was 0.82, and the AUC of combine 2 (CD4 + BMI + HB) was 0.831. To compare the predictive performances of combine 1 and combine 2, their C-Indexes were calculated, and the results were 0.806 (95% CI: 0.766, 0.846) and 0.798 (95% CI: 0.758, 0.839), indicating both models had a prediction accuracy of around 80%. Besides, no statistically significant difference in the C-Indexes between combine model was observed ($P=0.957$) (Fig. 4A). The discrimination between the two models was not large, but combine 2 involved fewer variables. Thus, combine 2 model was chosen and the three variables CD4, BMI and HB were preliminarily selected to construct a prediction model of three-year and five-year survival of PLHIV after ART.

Validation of prediction model

To verify the efficacy of the model in predicting the survival of PLHIV, bootstrap resampling was used for internal validation of the model. In the validation set, the AUCs of Shingles, CD4, BMI, HB and TC were 0.509, 0.821, 0.676, 0.77 and 0.654 in the ROC analysis chart (Fig. 3B).

The AUC of combine 1 achieved 0.802, and the AUC of combine 2 (prediction model) was also 0.802. The C-Indexes of combine 1 (0.786; 95% CI: 0.679, 0.893) and combine 2 (0.786; 95% CI: 0.681, 0.892) were similar and the difference was not statistically significant ($P=0.998$),

Table 2 Baseline demographics and clinical characteristics of patients in the training set and the validation set

Variable	Training set (N = 420)	Validation set (N = 180)	P value
Continuous variables			
Age, year	50.00 (40.00,57.00)	48.00 (37.75,61.00)	0.376
Interval from onset to diagnosis, day	52.00 (33.00,159.50)	46.50 (26.00,149.00)	0.107
ALT, U/L	21.65 (15.57,32.92)	22.15 (16.00,32.65)	0.629
AST, U/L	22.15 (17.90,30.33)	23.45 (18.95,31.23)	0.173
BMI, kg/m ²	21.85 (19.84,23.94)	21.48 (19.24,23.68)	0.106
CD4, cells/μL	259.50 (111.50,384.50)	243.50 (101.00,377.25)	0.79
TC, mmol/L	4.15 (3.55,4.82)	4.12 (3.48,4.80)	0.559
CR, μmol/L	69.00 (61.00,77.00)	69.00 (59.83,78.00)	0.778
FBG, mmol/L	5.60 (5.10,6.09)	5.37 (4.91,5.92)	0.051
HB, g/L	136.00 (119.75,151.00)	136.00 (117.75,152.25)	0.956
PLT, 10 ⁹ /L	172.00 (137.00,213.25)	167.50 (135.75,219.00)	0.808
TBIL, μmol/L	10.25 (7.80,14.00)	10.30 (6.80,13.40)	0.203
TG, mmol/L	1.38 (1.03,1.87)	1.32 (0.96,1.97)	0.458
WBC, 10 ⁹ /L	5.20 (4.10,6.38)	5.04 (4.01,6.38)	0.475
Discrete variables			
Gender, n(%)			
Male	386 (91.90)	160 (88.89)	0.276
Female	34 (8.10)	20 (11.11)	
Hepatitis B Virus, n(%)			
Negative	390 (92.86)	170 (94.44)	0.593
Positive	30 (7.14)	10 (5.56)	
Hepatitis C Virus, n(%)			
Negative	413 (98.33)	178 (98.89)	0.731
Positive	7 (1.67)	2 (1.11)	
Marital status, n(%)			
Unmarried	68 (16.19)	42 (23.33)	0.05
Married	352 (83.81)	138 (76.67)	
Shingles, n(%)			
No	375 (89.29)	165 (91.67)	0.458
Yes	45 (10.71)	15 (8.33)	
Infection route, n(%)			
Homosexual transmission	240 (57.14)	98 (54.44)	0.77
Heterosexual transmission	122 (29.05)	54 (30.00)	
Other transmission	58 (13.81)	28 (15.56)	
Baseline TB, n(%)			
No	407 (96.90)	175 (97.22)	1.00
Yes	13 (3.10)	5 (2.78)	
WHO clinical stage, n(%)			
I, II	189 (45.00)	75 (41.67)	0.655
III	102 (24.29)	43 (23.89)	
IV	129 (30.71)	62 (34.44)	
Continuous or intermittent fever, n(%)			
No	387 (92.14)	167 (92.78)	0.868
Yes	33 (7.86)	13 (7.22)	
Continuous diarrhea (> one month), n(%)			
No	397 (94.52)	165 (91.67)	0.202
Yes	23 (5.48)	15 (8.33)	

Abbreviations: ALT alanine aminotransferase, AST aspartate aminotransferase, BMI body mass index, CD4 CD4 cell count, WBC white blood cell, PLT blood platelet, HB hemoglobin, CR creatinine, TG triglyceride, TC total cholesterol, FBG fasting blood glucose, TBIL total bilirubin, TB Tuberculosis

Table 3 Univariable and multivariable Cox proportional hazards analysis of the training set

Variables	Univariable	P value	Multivariable	P value
	HR (95%CI)		AHR (95%CI)	
Female	1.364 (0.686,2.712)	0.376		
Marital Status (Married)	0.83 (0.496,1.388)	0.478		
Infection route				
Homosexual transmission	Ref	Ref	Ref	Ref
Heterosexual transmission	1.551 (0.984,2.446)	0.059#	1.327 (0.821, 2.145)	0.248
Other transmission	1.979 (1.142,3.428)	0.015*	1.193 (0.661, 2.154)	0.478
Baseline TB	2.024 (0.821,4.986)	0.125#	1.264 (0.491, 3.255)	0.584
HBV (Negative)	0.619 (0.227,1.689)	0.349		
HCV (Negative)	0.702 (0.098,5.04)	0.725		
Continuous diarrhea (> one month)	2.478 (1.286,4.775)	0.007*	2.271 (1.088, 4.742)	0.058
Continuous or intermittent fever	2.028 (1.141,3.605)	0.016*	0.905 (0.448, 1.680)	0.628
Shingles	1.905 (1.138,3.187)	0.014*	1.965 (1.140, 3.390)	0.018*
WHO clinical stage				
I, II	Ref	Ref	–	–
III	1.572 (0.908,2.721)	0.106#	–	–
IV	2.692 (1.678,4.318)	< .001***	–	–
Age, year	1.007 (0.99,1.023)	0.424		
Interval from onset to diagnosis, day	1.000 (1.000,1.000)	0.564		
CD4, cells/μL	0.995 (0.994,0.997)	< .001***	0.996 (0.995, 0.0.998)	< .001***
BMI, kg/m ²	0.795 (0.74,0.853)	< .001***	0.875 (0.809, 0.946)	< .001***
WBC, 10 ⁹ /L	0.969 (0.888,1.057)	0.476		
PLT, 10 ⁹ /L	0.998 (0.995,1.001)	0.254		
HB, g/L	0.968 (0.961,0.975)	< .001***	0.975 (0.965, 0.984)	< .001***
CR, μmol/L	1.002 (1,1.005)	0.099#	1.002 (0.999, 1.005)	0.175
TG, mmol/L	0.963 (0.849,1.093)	0.562		
TC, mmol/L	0.707 (0.57,0.877)	0.002*	0.821 (0.670, 1.006)	0.038*
FBG, mmol/L	0.731 (0.59,0.905)	0.004*	0.938 (0.769, 1.145)	0.442
AST, U/L	1.011 (1.006,1.017)	< .001***	1.001 (0.989, 1.014)	0.951
ALT, U/L	1.008 (1.002,1.014)	0.008*	1.002 (0.992, 1.012)	0.647
TBIL, mmol/L	0.991 (0.957,1.027)	0.617		

Note: # $p < 0.2$; * $p < 0.05$; *** $p < 0.001$

Abbreviations: CD4 CD4 cell count, WBC white blood cell, PLT blood platelet, HB hemoglobin, CR creatinine, TG triglyceride, TC total cholesterol, FBG fasting blood glucose, AST aspartate aminotransferase, ALT alanine aminotransferase, TBIL total bilirubin, TB Tuberculosis

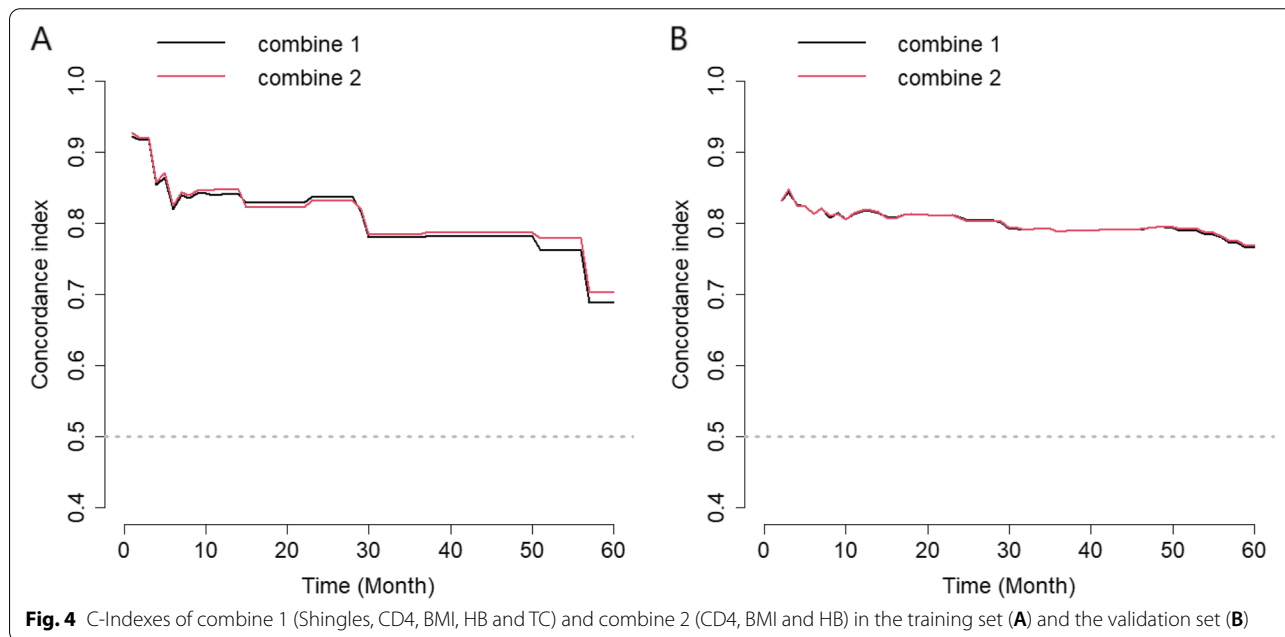
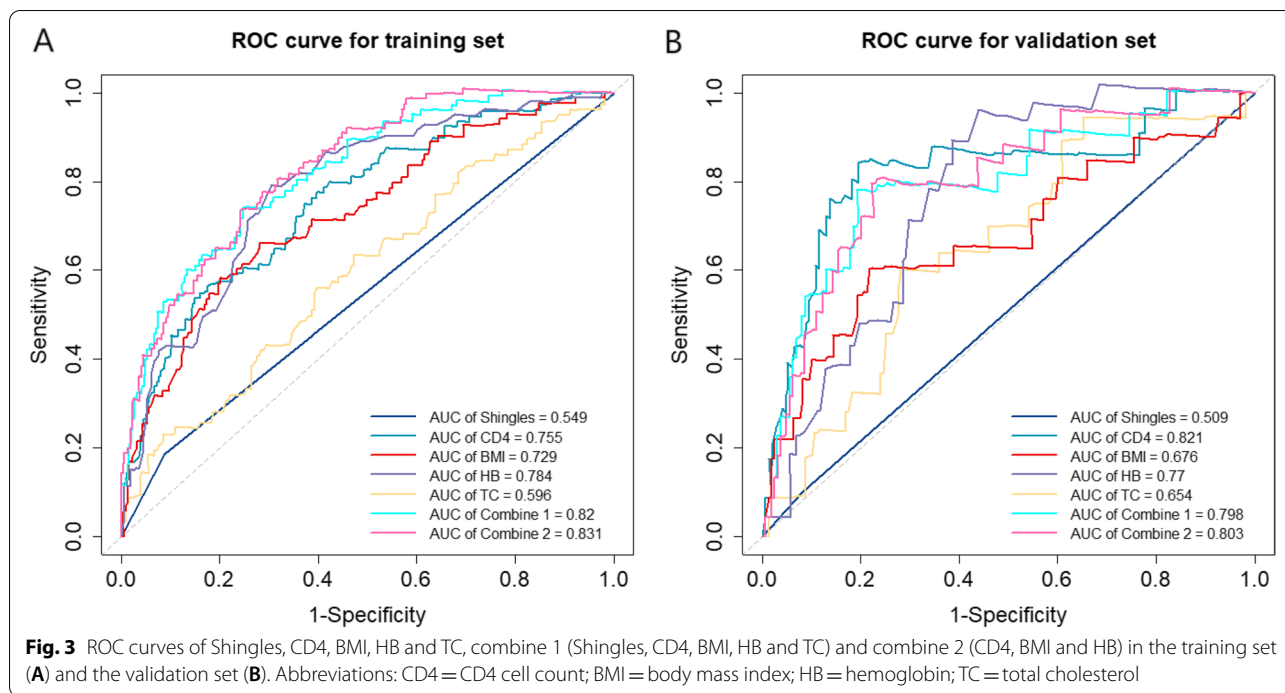
which showed that the discrimination of combine 1 and combine 2 (prediction model) was not very large (Fig. 4B). The calibration curve also exhibited a high consistency in predicting the survival of PLHIV (especially in the first 3 years after ART initiation) (Fig. 5).

As shown in Fig. 6, in both the training set and the validation set, the prediction model (combine 2) showed better performance. Overall, the DCA curve demonstrated that the prediction model (combine 2) could make valuable and profitable judgements. In addition, among the detected factors, CD4 was more beneficial than the other routine clinical laboratory indicators in predicting the three-year and five-year survival probabilities of PLHIV. In Fig. 6D, DCA curve showed that the prediction model

had no good benefits in predicting five-year survival probabilities of PLHIV in the validation set.

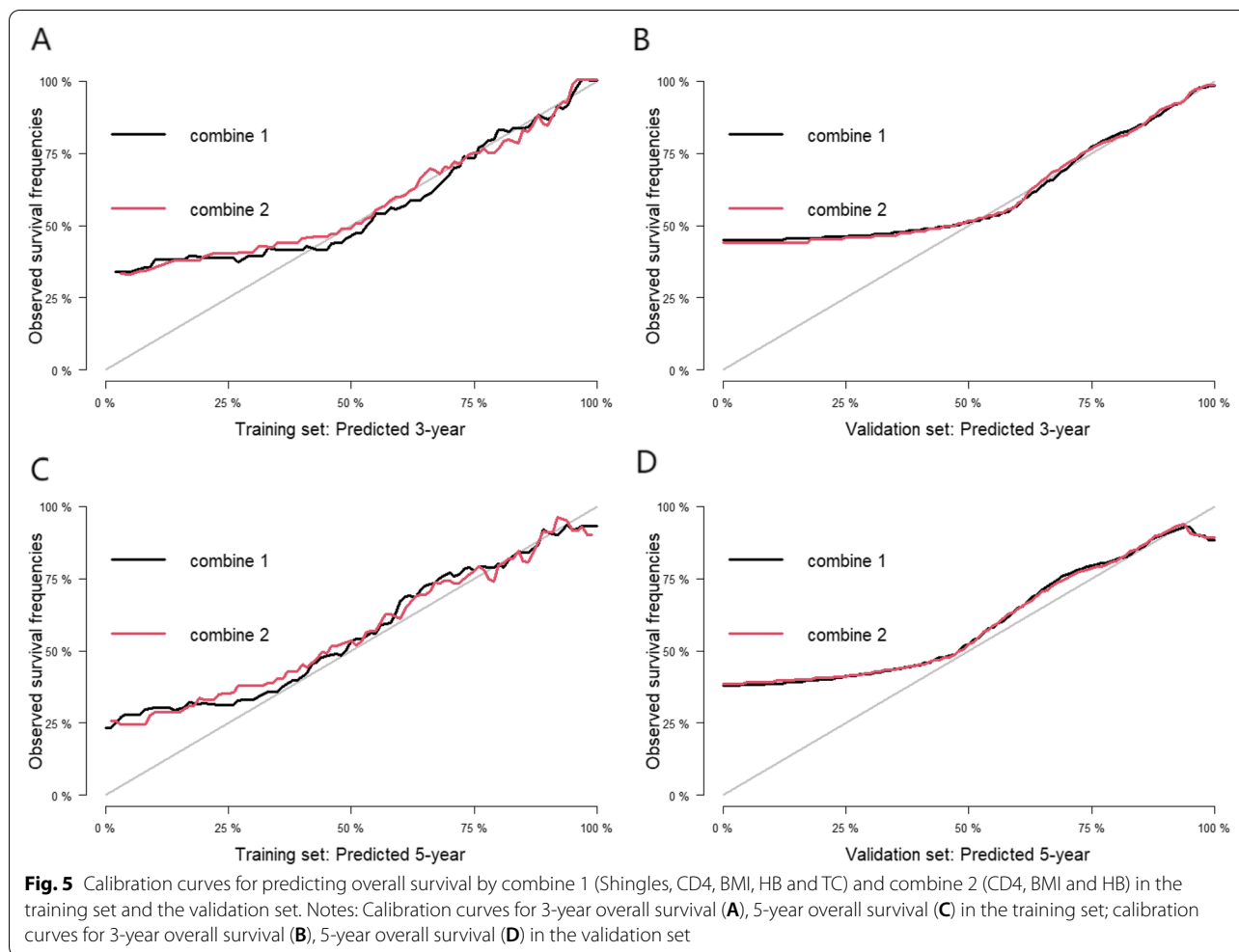
Performance of nomogram

A nomogram was drawn according to the determined prediction model. As seen in Fig. 7, each selected predictor was assigned with a score according to its value in the nomogram based on the established prediction model. Then a vertical line perpendicular to the Point axis was drawn from this point. The intersection point on the Point axis represented the score under the determined value of the predictor. For example, when CD4 was 1200 cells/μL, the score was 0 point; when



BMI was 12kg/m², the score was 63 points. By analogy, the score of each predictor could be determined, and summed up. Similarly, after the total score was calculated, a vertical line was drawn from the point of the patient’s total score on the Total Points axis to the axis of survival probability (such as three-year survival

probability or five-year survival probability). The intersection point on the axis of survival probability represented the patient’s three-year or five-year survival probability.



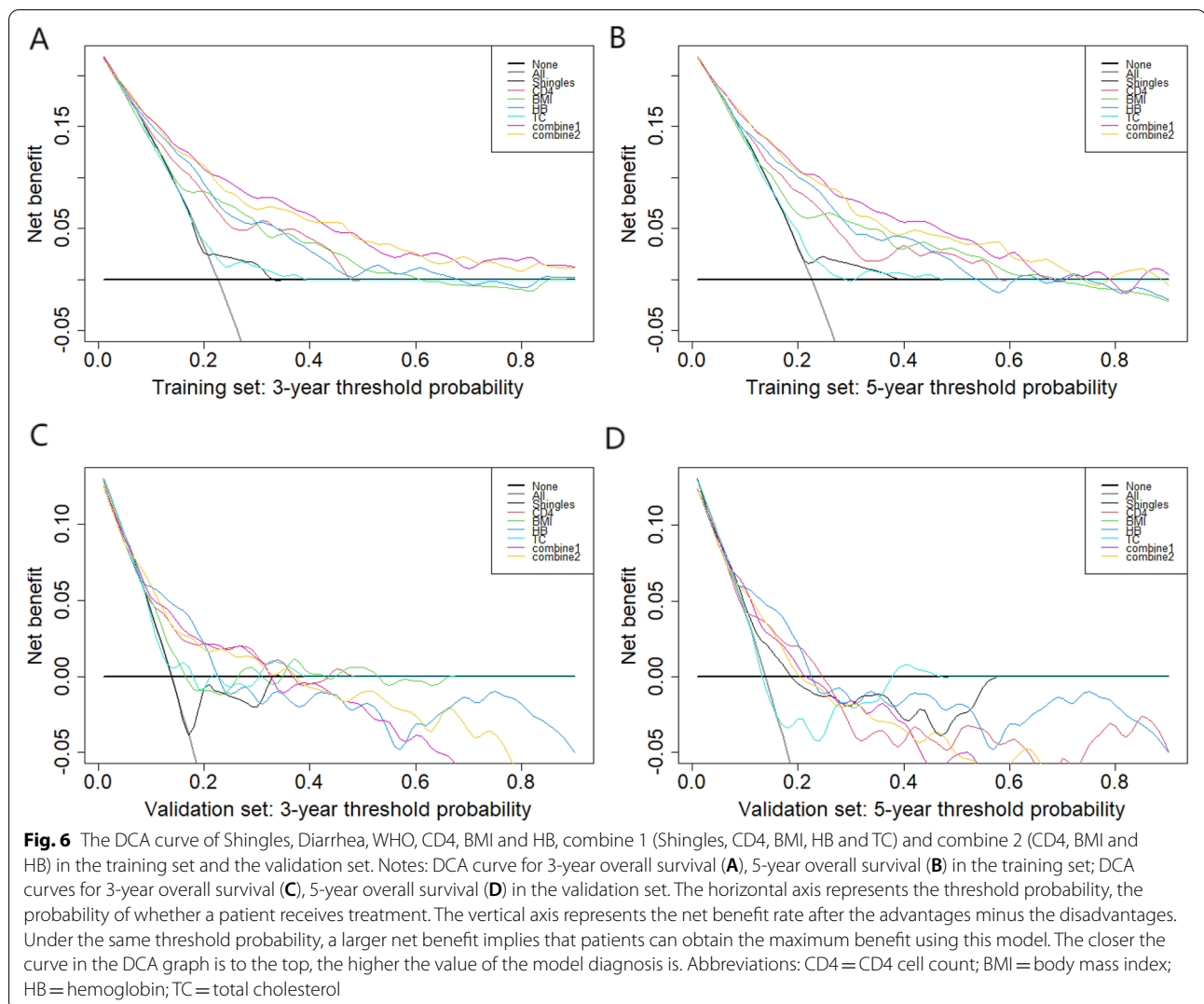
Discussion

Although the survival of PLHIV has been improved significantly with the promotion of free ART, a rapid and accurate prediction can benefit the personalized management of PLHIV and the allocation of medical resources [26].

For prognosis, due to a longitudinal temporal logic between predictors and outcome, the cohort study is used to analyze the data and fit a prognostic model. Randomized controlled clinical trials are considered as a prospective cohort study with more rigorous inclusion criteria, which therefore can be used to establish a prognostic model. However, it has limitations in extrapolation. Due to the population selection bias and information bias, retrospective cohort studies are not suitable for constructing a prognostic model, while nested case-control or case cohort studies are more economical and feasible for studies with rare outcomes or expensive predictive factor measurements. To decrease the influence of the limitation, we took into account the survival time

when performed PSM. Based on this nested case-control study of an HIV/AIDS ART cohort in Nanjing, the relationship between routine laboratory indicators and the survival probability of PLHIV was evaluated. A prognostic model (including CD4, BMI and HB) with satisfactory discrimination and calibration was developed to predict the three-year and five-year survival of PLHIV receiving ART. Then the result of this prognostic model was shown in the form of a nomogram.

Nomogram is simple, direct and effective in predicting the prognosis of PLHIV [24]. In this study, the multivariable Cox proportional hazards regression model indicated that the five factors (Shingles, CD4, BMI, HB and TC) were associated with the HIV/AIDS-related survival time. To overcome the limitation of a single predictor and simplify the prediction procedure, three detected factors (CD4, BMI and HB) were combined to construct a prognostic model to predict the three-year and five-year survival of ART-treated PLHIV, which exhibited a high consistency.



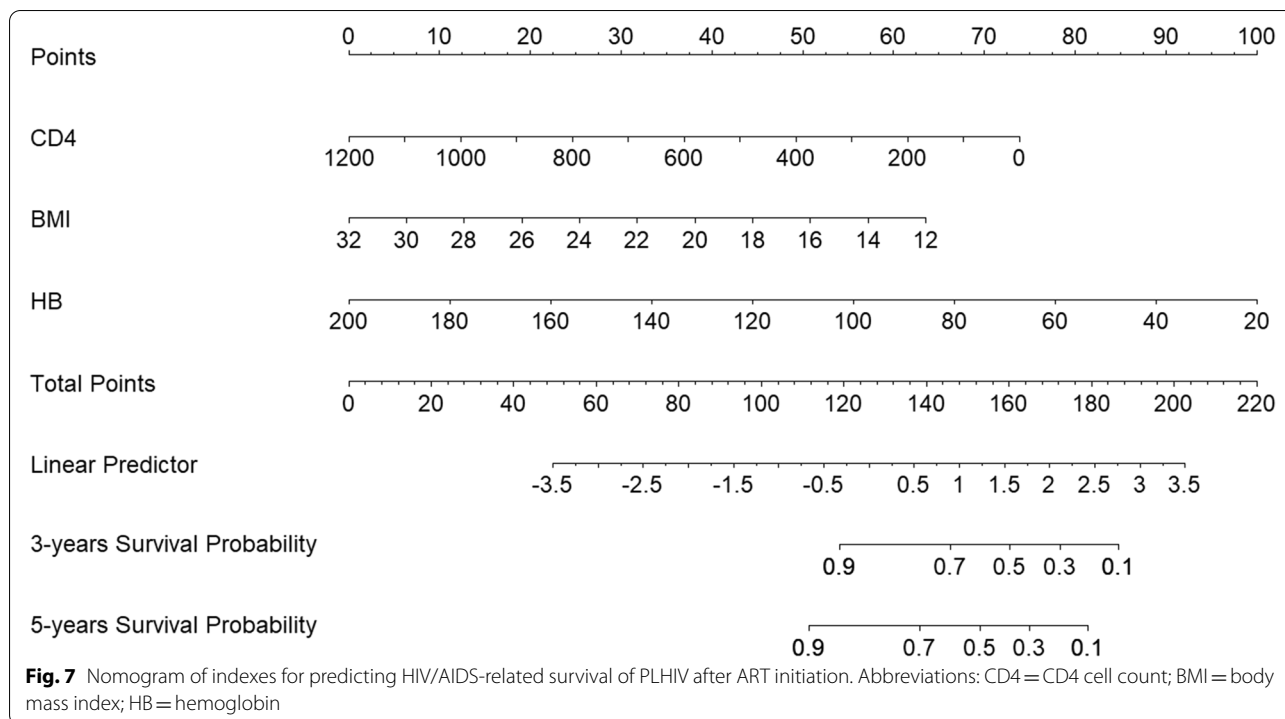
WHO clinical stage had a close association with PLHIV survival [13] but was excluded in the nomogram. The main reason was that there was a strong relationship between WHO clinical stage and CD4 in the current study, which caused multicollinearity. In addition, the laboratory indicators (CD4) usually are more sensitive in predicting survival rate of PLHIV than the clinical indicators (WHO clinical stage). In recent years, many researchers have reported that some laboratory indicators are connected with the survival of PLHIV. In this study, CD4, BMI and HB were significantly correlated with the survival of PLHIV and showed good consistency with these published studies [10, 16, 21, 26].

An obesity paradox was seen in this predictive nomogram of PLHIV, and those with high BMI had a low risk of death. This may be due to the fact that the protective effect of BMI helps preserve the immune system response and slow the progression of HIV [31]. There is some evidence

that a higher BMI is associated with more robust CD4 recovery in ART-treated patients [32]. Previous studies also suggested that the immune reconstitution on ART was often the highest among overweighted patients [33].

DCA is commonly applied to assess the efficacy of specific clinical prediction models [34]. In this study, DCA was used to assess the potential clinical benefits of nomogram, which revealed that nomogram was more effective and accurate than a single indicator in forecasting the survival of PLHIV. Prediction models are always less powerful in predicting outcomes during a long time. With more samples in the future, the performance of prediction models might be improved.

The present model has a limitation. It was established based on a few easily collected and low-cost predictors due to the underdeveloped technology in the past. However, as the economy and technology evolve, clinical prediction models that involve a larger number of data (big



data) will be developed. Hopefully, more complex models and algorithms based on machine learning and artificial intelligence will provide more benefits to medical workers, PLHIV and medical decision makers.

Abbreviations

CD4: CD4 cell count; WBC: White blood cell; PLT: Blood platelet; HB: Hemoglobin; CR: Creatinine; TG: Triglyceride; TC: Total cholesterol; FBG: Fasting blood glucose; AST: Aspartate aminotransferase; ALT: Alanine aminotransferase; TBIL: Total bilirubin; TB: Tuberculosis; BMI: Body mass index; CD8: CD8 cell count; VL: Viral load; PSM: Propensity-score matching; ART: Antiretroviral therapy; PLHIV: People living with HIV; C-Index: Concordance index; DCA: Decision curve analysis.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12889-021-12249-8>.

Additional file 1: Supplementary data. Raw data of HIV/AIDS patients used and analyzed in current study. Note: Variable names and values are described as follows. Infection route (1 = Homosexual transmission, 2 = Heterosexual transmission, 3 = Other transmission); Gender (1 = Male, 2 = Female); Marital status (1 = Unmarried, 2 = Married); Art = time of initiating antiretroviral therapy; TB = tuberculosis (0 = No, 1 = Yes); Continuous diarrhea (0 = No, 1 = Yes); Continuous or intermittent fever (0 = No, 1 = Yes); Shingles (0 = No, 1 = Yes); WHO clinical stage (1 = stage I or II, 2 = stage III, 3 = stage IV); ALT = alanine aminotransferase; AST = aspartate aminotransferase; BMI = body mass index; CD4 = CD4 cell count; WBC = white blood cell; PLT = blood platelet; HB = hemoglobin; CR = creatinine; TG = triglyceride; TC = total cholesterol; FBG = fasting blood glucose; TBIL = total bilirubin; HBV = Hepatitis B Virus (0 = Negative, 1 = Positive); HCV = Hepatitis C Virus (0 = Negative, 1 = Positive); Status = survival status at the last follow-up Status (0 = alive, 1 = dead); End = observation end point.

Acknowledgements

The authors thank all the study participants for their contributions and the staff at all participating institutions for their support.

Authors' contributions

GF, LW⁶, and ZL curated and provided data. ZL and JX screened the data. LL and GF performed laboratory tests. FJ, YX, LL, KW and LW⁴ performed data analysis. FJ and YX drafted the manuscript. NW and HX provided guidance for epidemiological analysis throughout the study. ZZ and ZP adjusted the research framework, provided financial support, and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported in part by the National Natural Science Foundation of China (82073673, 91846302), Nanjing key medical science and technology development projects (ZKX19050), and the National S&T Major Project Foundation of China (2017ZX10201101, 2018ZX10715002).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The data were extracted from the Nanjing AIDS Prevention and Control Information System (AIDS-PCIS), which was established by the China Center for Disease Control and Prevention (CCDC). All the methods carried out in our study are accorded with relevant guidelines. The AIDS-PCIS protocol was approved by the institutional review boards at the CCDC. Informed consent was obtained from the subjects before their enrollments. The ethical approval for the study was also obtained from the Ethics Review Board of Nanjing Center for Disease Control and Prevention and the ethical committee of Nanjing Medical University ("F", "CH", "Nanjing Med U", "FWA00001501", "NANJING", 11/21/2004). I have read and have abided by the statement of ethical standards for manuscripts submitted to BMC Public Health.

Consent for publication

Not applicable.

Competing interests

All authors declare that they have no conflict of interest or financial conflicts to disclose.

Author details

¹Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 211166, Jiangsu, China. ²Department of AIDS and STDs control and prevention, Nanjing Municipal Center for Disease Control and Prevention, Nanjing 210003, Jiangsu, China. ³Jiangsu Provincial Key Laboratory of Geriatrics, Department of Geriatrics, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China. ⁴National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China. ⁵Department of STDs/AIDS Prevention and Control, Jiangsu Center for Disease Prevention and Control, Jiangsu 210027, China. ⁶Division of Infectious Diseases, Key Laboratory of Surveillance and Early-warning on Infectious Disease, Chinese Center for Disease Control and Prevention, Beijing 102206, China. ⁷Department of Epidemiology and Biostatistics, School of Public Health, China Medical University, Beijing 110001, China.

Received: 1 February 2021 Accepted: 8 November 2021

Published online: 06 January 2022

References

- Zayeri F, Ghane ET, Borumandnia N. Assessing the trend of HIV/AIDS mortality rate in Asia and North Africa: an application of latent growth models. *Epidemiol Infect.* 2016;144(3):548–55.
- Zhang FJ, Jennifer P, Lan Y, et al. Current progress of China's free ART program. *Cell Res.* 2005;15(11):877–82.
- Bansi L, Gazzard B, Post F, et al. Biomarkers to monitor safety in people on art and risk of mortality. *J Acquir Immune Defic Syndr.* 2012;60(1):51–8.
- Wang CW, Chan CLW, Ho RTH. HIV/AIDS-related deaths in China, 2000–2012. *AIDS Care.* 2015;27(7):849–54.
- Ning SP, Xue ZD, Wei J, et al. HIV/AIDS related mortality in southern Shanxi province and its risk factors. *Chin J Epidemiol.* 2015;36(3):245–9.
- Yoshikura H. Shift of HIV/AIDS deaths to an older age and gender difference: inferences derived from the vital statistics of Japan. *Jpn J Infect Dis.* 2019;72(6):359–67.
- Tang HL, Hou L, Han J, Li J, et al. Effects of standardized follow-up program among newly diagnosed HIV/AIDS cases in 2010. *Chin J Epidemiol.* 2016;37(12):1602–7.
- Li Y, Wang J, He SF, et al. Survival time of HIV/AIDS cases and related factors in Beijing, 1995–2015. *Chin J Epidemiol.* 2017;38(11):1509–13.
- Zeng YL, Tang HL, Li JM, et al. Survival analysis of people living with HIV/AIDS in Sichuan province, 1991–2017. *Chin J Epidemiol.* 2019;40(3):309–14.
- Zhang G, Gong Y, Wang Q, et al. Outcomes and factors associated with survival of patients with HIV/AIDS initiating antiretroviral treatment in Liangshan Prefecture, southwest of China: A retrospective cohort study from 2005 to 2013. *Medicine (Baltimore).* 2016;95(27):e3969.
- Zhang N, Zhu XY, Wang GY, et al. Survival status and influencing factors of HIV/AIDS on highly active anti-retroviral therapy in Shandong province. *Chin J Epidemiol.* 2019;40(1):74–8.
- Justice AC, Modur S, Tate JP, et al. Predictive accuracy of the veterans aging cohort study (VACS) index for mortality with HIV infection: a north American cross cohort analysis. *J Acquir Immune Defic Syndr (1999).* 2013;62(2):149.
- Silverman RA, John-Stewart GC, Beck IA, et al. Predictors of mortality within the first year of initiating antiretroviral therapy in urban and rural Kenya: a prospective cohort study. *PLoS One.* 2019;14(10):e0223411.
- Seyoum D, Degryse JM, Kifle YG, et al. Risk factors for mortality among adult HIV/AIDS patients following antiretroviral therapy in southwestern Ethiopia: an assessment through survival models. *Int J Environ Res Public Health.* 2017;14(3):296.
- Misgina KH, Weldu MG, Gebremariam TH, et al. Predictors of mortality among adult people living with HIV/AIDS on antiretroviral therapy at Suhul hospital, Tigray, northern Ethiopia: a retrospective follow-up study. *J Health Popul Nutr.* 2019;38(1):37.
- Jiang J, Qin X, Liu H, et al. An optimal BMI range associated with a lower risk of mortality among HIV-infected adults initiating antiretroviral therapy in Guangxi, China. *Sci Rep.* 2019;9(1):1–10.
- Aziz N, Quint JJ, Breen EC, et al. 30-year longitudinal study of hematological parameters of HIV-1 negative men participating in Los Angeles multicenter AIDS cohort study (MACS). *Lab Med.* 2019;50(1):64–72.
- Harris RJ, Sterne JAC, Abgrall S, et al. Prognostic importance of anaemia in HIV-1 infected patients starting antiretroviral therapy: collaborative analysis of prospective cohort studies in industrialized countries. *Antivir Ther.* 2008;13(8):959.
- Belperio PS, Rhew DC. Prevalence and outcomes of anemia in individuals with human immunodeficiency virus: a systematic review of the literature. *Am J Med.* 2004;116(7):27–43.
- Camon S, Quiros C, Saubi N, et al. Full blood count values as a predictor of poor outcome of pneumonia among HIV-infected patients. *BMC Infect Dis.* 2018;18(1):189.
- Bisson GP, Ramchandani R, Miyahara S, et al. Risk factors for early mortality on antiretroviral therapy in advanced HIV-infected adults. *AIDS (London, England).* 2017;31(16):2217.
- Gardner LI, Holmberg SD, Williamson JM, et al. Development of proteinuria or elevated serum creatinine and mortality in HIV-infected women. *J Acquir Immune Defic Syndr.* 2003;32(2):203–9.
- Driver TH, Scherzer R, Peralta CA, et al. Comparisons of creatinine and cystatin C for detection of kidney disease and prediction of all-cause mortality in HIV-infected women. *AIDS (London, England).* 2013;27(14):2291.
- Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162(1):W1–73.
- Park SY. Nomogram: an analogue tool to deliver digital knowledge. *J Thorac Cardiovasc Surg.* 2018;155(4):1793.
- McNairy ML, Jannat-Khah D, Pape JW, et al. Predicting death and lost to follow-up among adults initiating antiretroviral therapy in resource-limited settings: derivation and external validation of a risk score in Haiti. *PLoS One.* 2018;13(8):e0201945.
- Hou X, Wang D, Zuo J, et al. Development and validation of a prognostic nomogram for HIV/AIDS patients who underwent antiretroviral therapy: data from a China population-based cohort. *EBioMedicine.* 2019;48:414–24.
- Benedetto U, Head SJ, Angelini GD, et al. Statistical primer: propensity score matching and its alternatives. *Eur J Cardiothorac Surg.* 2018;53(6):1112–7.
- Van Den Berg HA. Occam's razor: from Ockham's via moderna to modern data science. *Sci Prog.* 2018;101(3):261–72.
- Efron B. Bootstrap methods: another look at the jackknife. *Ann Stat.* 1979;7(1):1–26.
- Shor-Posner G, Campa A, Zhang G, et al. When obesity is desirable: a longitudinal study of the Miami HIV-1-infected drug abusers (MIDAS) cohort. *J Acquir Immune Defic Syndr (1999).* 2000;23(1):81–8.
- Koethe JR, Jenkins CA, Lau B, et al. Higher time-updated body mass index: association with improved CD4+ cell recovery on HIV treatment. *J Acquir Immune Defic Syndr (1999).* 2016;73(2):197.
- Koethe JR, Jenkins CA, Shepherd BE, et al. An optimal body mass index range associated with improved immune reconstitution among HIV-infected adults initiating antiretroviral therapy. *Clin Infect Dis.* 2011;53(9):952–60.
- Kerr KF, Brown MD, Zhu K, et al. Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. *J Clin Oncol.* 2016;34(21):2534.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.